

Analytical Information Markup Language (AnIML): A Tool to Insure the Usability of Multiple Format Scientific Data Over Time

The way scientific data is stored and analyzed changes with the advent of new computer technologies. NIST scientists are developing Analytical Information Markup Language (AnIML) that will enable analytical data to be interchanged over the web, converted to other formats, validated, or visualized in multiple formats using existing XML-based tools. Although AnIML is being developed for chemical spectroscopy and chromatography data, it is capable of describing almost any data. For the first time, scientists will be able to seamlessly compare data from different instruments, applications and databases.

G. W. Kramer, P. H. Gleichmann, and K. S. Placzek (Div. 831), R. Schaefer, A. Roth, and R. Jopp (Fachhochschule Wiesbaden) and P. J. Linstrom (Div. 838)

Interchanging analytical data and their associated "scientific metadata" across space and time, from application-to-application, and to/from applications and databases has often been hampered by multiple, proprietary, incompatible data formats. Over the past 15 years, chemical spectroscopy and chromatography data have been transferred using pre-web-based technologies in conjunction with either the ASTM Analytical Data Interchange (ANDI) standards or the IUPAC JCAMP-DX (Joint Committee on Atomic and Molecular Physics Data Exchange) standards. While these procedures allow the interchange data from instrument to instrument, data interchange from instrument to application (e.g., importing data from an instrument into an Excel spreadsheet), from application to application, or from application to and from databases is not as well supported. Modern laboratory management concepts such as electronic laboratory notebooks require simple, common mechanisms for interchanging data between instruments, applications, and databases.

NIST is developing new computer language to enable the analysis and usability of multiple forms of data, regardless of its original form, storage media or application. This is critical for archiving irreplaceable data needed in all biological fields.

The rapid pace of information technology and computing hardware innovation has exacerbated the problem of ar-

chiving of scientific data. Analytical information stored on early digital media (for example, 8-inch floppy disks) 20 years ago may be less accessible today than such information stored in a paper format 20 years prior. The difficulty is not solely being able to still read the data on the storage media, but it is also being able to decipher the proprietary formats used by obsolete applications and operating systems.

<  AnIML >
<Analytical Information Markup Language=
"For Spectroscopy & Chromatography Data" >
< /  AnIML >

In an effort aimed to solve the interchange and archiving dilemma for analytical chemistry data, ASTM Subcommittee E13.15 on Analytical Data is creating the Analytical Information Markup Language (AnIML) for describing chromatography and spectroscopy data and their metadata. Based on XML (eXtensible Markup Language) and its associated technologies, AnIML facilitates access to analytical data by building in descriptions of the data and metadata with delimited tags in the same way that HTML (HyperText Markup Language) describes the display of items on a webpage. AnIML is built around a core schema that defines ways for describing almost any data. A separate Technique schema describes Technique Definition files that are used to constrain the data description mechanisms in the core for a given analytical technique to those commonly accepted, to delineate the metadata items commonly associated with such domain data, and to permit content extension by vendors and users without changing the core schema.

Once in AnIML format, analytical data can be interchanged over the web, converted to other formats, validated, or visualized in multiple formats using existing XML-based tools. The "self-describing" XML-based format used in AnIML allows applications to utilize the data easily based on the tags that describe each data entry. This affords the possibility for creating generic programs that can visualize, display, and print AnIML-based spectroscopic and chromatography data from any source. Using AnIML formats, casual users would no longer have to have access to or have to learn proprietary vendor software just to look at analytical results. Since data formats are open and self-documenting in XML and since most XML

is text-based, data stored in AnIML format should be accessible for as long as the media it is stored on can be read. Furthermore, AnIML ensures the integrity of the data through the use of digital signatures and provides for the data tracking, verification, and validation necessary for use in regulated industries.

The AnIML effort is international with participants from Germany, France, the United Kingdom, and Jamaica. Most of the participants are from instrument manufacturers or the pharmaceutical sector, e.g., Agilent, GlaxoSmith-Kline, Waters, Shimadzu, the TopCombi Group, Jasco, ADC Labs, Thermo Electron Corp., Bristol-Myers Squibb. Further information about AnIML can be found at <http://animl.sourceforge.net>.

Publications:

A. Roth, R. Jopp, R. Schäfer, and G.W. Kramer, **Automated Generation of AnIML Documents by Analytical Instruments**, *Journal of the Association for Laboratory Automation*, 2006, **11**, 247.